DOCUMENT RESUME

ED 410 265 TM 027 047

AUTHOR Roeber, Edward D.
TITLE Designing Coordinated Assessment Systems for IASA Title I.
PUB DATE Mar 97
NOTE 28p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Compensatory Education; *Coordination; *Educational Assessment; Elementary Secondary Education; Federal Aid; Federal Legislation; Program Evaluation; *Standards; *Test Construction; Test Format; Test Use
IDENTIFIERS *Improving Americas Schools Act 1994 Title I

ABSTRACT

This paper describes an approach that states and local districts might use to develop a coordinated assessment system so that information collected at state and local levels about programs under Title I of the Improving America's Schools Act (IASA) would be complementary and not redundant. The reauthorization of Title I provides a unique opportunity for states and districts to work together to develop the assessments mandated for Title I accountability. A coordinated set of assessments must be designed so that assessments used at different levels of the educational system fit together to provide a more complete picture of student achievement than one assessment could provide alone. It will be essential to consider the purposes of the assessment and the achievement information different users of the assessment will need. An example of a coordinated assessment system is presented, emphasizing what the state must develop and what the local district must do. An approach that is both top-down and bottom-up must be developed to reflect the various audiences for assessments. Steps are presented to result in a cross-level coordinated assessment system. These include: (1) state and local standards coordination; (2) development of the assessment blueprint; and (3) developing performance standards. Two attachments describe the purposes of assessment and its formats. (Contains two figures.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 410 265

Designing Coordinated Assessment Systems for IASA Title I

State Collaborative on Assessment and Student Standards
Comprehensive Assessment Design for IASA Title I/Goals 2000

March, 1997

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Edward Roeber
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Edward D. Roeber
Director, Student Assessment Programs
Council of Chief State School Officers

TM027047

BEST COPY AVAILABLE

Designing Coordinated Assessment Systems for IASA Title I

Introduction

The purpose of this paper is to describe an approach that states and local districts might use to develop a coordinated assessment system so that the information collected at the state and local levels are complementary, not redundant. Since many states have developed or are developing content and performance standards, as well as assessments, this paper can also show how to retrofit coordination of state and local standards and assessments onto existing or developing standards and assessment systems developed separately at the state and local levels. The advice suggested in this paper can serve as a model for states and their local districts to strive to attain, while using existing standards and assessments. The reauthorization of Title I provides a unique opportunity for states and local districts to work together to develop assessments needed at different levels of the educational system so that they provide a more complete picture of student performance and are used to help Title I (and all other) students learn at high levels.

Title I Assessment Requirements

Each state must develop or adopt student assessments to be administered annually to students in at least one grade at the elementary, middle school, and high school levels, and in at least the subjects of mathematics and reading or language arts. The measures used, and the standards they are based on, must be the same standards and assessments required of all students in the state. States are given considerable flexibility to determine what the content and performance standards will look like, how they are developed and how they are implemented. States have until the 1997-98 school year in which to develop and implement the content and performance standards.

States have until 1999-2000 develop, pilot, and implement the assessments of students. These assessments will be the primary means for determining whether districts and schools receiving Title I funds have effectively served students. For Title I purposes, the assessments must:

- be administered at some time during grades 3 through 5, grades 6 through 9, and grades 10 through 12;
- include multiple measures of the state's content standards;
- include measures of higher-order skills;
- provide accommodations for students with diverse learning needs; and
- include limited English proficient students, in the language most likely to yield accurate and reliable information about them in subjects other than English.

This paper will define coordinated assessment systems, describe some reasons why such coordination is both feasible and desirable, and suggest ways in which coordinated assessment systems could be created. It will suggest strategies by which states could work to create such systems from the outset. Of course, many states and local districts have systems already in place that were not created in the manner described in this paper. It is hoped that even in these cases that the ideas presented in this paper will serve to encourage states and local districts to look

for ways of achieving partial coordination, perhaps through retro-fitting standards or assessments across levels.

Chapter Two

Coordinated Assessment Design – An Introduction

There are several basic elements that will need to be determined as the coordinated assessment plan is being developed. First, what is meant by a "coordinated" assessment system? Second, what are the purposes for assessment, and can these be met with one assessment? Are the purposes for assessment the same at the state and local levels? If not, what are the information needs at different levels? How do the types of assessment exercises to be used fit into the coordinated assessment system? When starting from scratch, states and local districts will have different options than when one or both levels has already developed standards and assessments and coordination of the results is desired. This chapter will present both the ideal, start-from-scratch model, as well as suggestions about how this ideal model can be approximated when retrofitting the ideal onto existing standards and assessments.

Coordinated, Comprehensive Assessment Systems

When designing state and local student assessments, an essential element that should occur is how the assessment program will "fit" together with other assessments in use or development to make up a *coordinated*, comprehensive system of assessments. This is important because it is usually difficult or impossible to use one assessment instrument for multiple purposes (whether it is one program at the state level or a program implemented by local school districts); none of the purposes may be done well.

What is meant by the terms "coordinated" and "system?" A coordinated set of assessments is defined as a set of assessments used at different levels of the educational system that fit together with one another to provide a more complete picture of student achievement than any one assessment could provide alone. A system of assessments means a group of assessments, of a different nature or type, selected to complement one another. Duplicative data collection (the collection of the same information using the same type of assessments) is eliminated, yet multiple sources of information are used, especially when important decisions are being made about students, schools or instructional programs. Different formats for assessment are used to create a composite of student and school information to better inform the user.

State and local assessment systems could be built in tandem, based on a common set of content standards so that the skills assessed are related, with different assessment tools used at different levels (each well suited for its intended purpose), and the assessments could work together to provide a more complete and coordinated picture of student and school program performance. When assessment systems have already been developed at the state and/or local levels, coordination could occur in one of two ways: the assessment system at one level (e.g., the state level) is used by the other level (e.g., local school districts) as the basis of a revision of this work. Alternatively, the two levels could look for commonalties among the standards and assessments at the two levels and strive to report information derived from assessments of these "common" standards. Either approach begins the process of coordination, which can be furthered once the standards at one or both levels are revised in the future.

Rationale for a Coordinated Assessment System

Coordinated assessment *systems* make sense because they may reduce the collection of unintentionally redundant information, use resources available for assessment to collect information most useful for the decisions that need to be made at each educational level, and serve to reduce the number of "mixed messages" that local educators and the public receive about "what is important." By developing one set of content standards, with appropriate curricula and instructional strategies, the likelihood that students are taught the important skills is also increased. When retrofitting state and local level assessments, the participants would need to look for existing commonalties among the standards at the two levels, so that common assessment information and instruction can be emphasized.

Since states and even local school districts have little ability to dictate what gets taught or learned in individual classrooms, a current system of uncoordinated standards and assessments merely increases the probability that different students will learn different knowledge and skills, and that perhaps the students with the greatest educational needs, the students being served by IASA Title I, will not be taught content standards as rigorous as those with other students.

Purposes for Assessment

There are several purposes for large-scale assessment at the state or local levels. For example, student assessment is viewed as the means of setting higher, more rigorous standards for student learning, focusing staff development efforts for the nation's teachers, encouraging curriculum reform and improving instruction and instructional materials in a variety of subject matters and disciplines. Assessment may also serve to hold schools accountable for whether these reforms have occurred and have been effective. Large-scale assessment has been a key policy tool used to attempt to affect change in the nation's schools. However, what we have learned over the past decade is that assessment programs that feature accountability for performance as a key purpose are often unable to fulfill the equally popular purpose of improving instruction. This is not only because accountability measures administered at the state level tend not to provide detailed information to teachers on a timely basis; the information often does not assess students in a fashion most related to day-to-day instruction.

One response to these issues regarding statewide assessments has been to develop teacher-based assessments in the hope that, while they provide more valuable information to teachers, they could also be used to aggregate upwards to provide system accountability information at the school, district, and state levels. While this is possible to do, and there have been a couple of instances where this is planned or has taken place, the types of assessments that are most useful to teachers do not often lend themselves to the public credibility demanded of accountability assessments; they are often developed, administered and scored by individual classroom teachers. Concerns have been raised about whether the pressure of accountability would corrupt the trustworthiness and value of the information to teachers. Clearly, the purposes for assessment may differ at different levels of the educational system.

Another way to view this is depicted in Figure 1. As Figure 1 shows, there are

purposes for assessment that are much more likely to occur at the national or state levels than at the classroom or schools levels. Monitoring school and district performance is a purpose often ascribed to state assessment programs. Another equally valued purpose is the improvement of individual student's educational programs. Because state assessment programs can efficiently collect information across all students at a grade level, they are ideally suited to fulfill the accountability or monitoring purpose. Yet, while any data set could be used to help individual students learn, it is less likely that such information will be helpful to classroom teachers than information collected much closer to the classroom.

Figure 1
Assessment Purposes at Different Levels of Education

	Purposes for Assessment						
Education Level	Monitor	Account.	Improve Student Perf.	Allocate Resources	Selection / Placement	Certification	Program Evaluatio
National	XXX	X		X		X	
State	XXX	XXX	X	XX		XXX	X
District	XX	XXX	XX	XX	XX	XX	XX
School	X	XX	XXX	X	XXX	X	XX
Classroom		X	XXX		XX	X	X
Student		X	XXX		X	X	X

Note: Each X indicates the levels at which each type of assessment might be used most efficiently at each level. Multiple X's indicate the strength of the appropriateness of assessment purpose at each educational level.

Hence, if the purpose for assessment is to help individual students learn better, some type of assessment established at the local level is needed, while if the purpose is to hold schools accountable for improving all students' performance, then some form of statewide assessment is needed. One practical implication of this idea is that a statewide assessment program may not be the most efficient or effective means of accomplishing all assessment information needs. Nor is it likely that classroom assessment, as valuable as it is for accomplishing certain needs, will be effective in meeting all assessment purposes. For the assessment system not to send disjointed, conflicting messages, however, the assessments used at these different levels ought to be measuring the same or related content standards. A more complete description of the various purposes for assessment is contained in Attachment A.

Achievement Information Needed at Different Levels

It is important to consider the information about student achievement needed by different individuals. Certainly, the information needs of the parent are different from those of the teacher; the parent wants to know what the student can do and not do, while the teacher is more concerned with what are the

student's strengths (what aspects of the learning task has the student accomplished) and what additional work is needed. These information needs are different from those of the building principal, who needs to know if achievement in the school is comparable to that elsewhere and more broadly, whether students are learning what they need to learn.

At the district level, the concern may be more whether the achievement needs are greater in mathematics than reading, so that additional resources (extra staff and additional instructional materials) should be allocated for mathematics instruction.

At the state level, the concern is often whether there is equity in school programs and whether differences in student performance are due to the lack of resources in some schools; issues of funding, incentive programs for improving in key areas such as mathematics, staff training and so forth are discussed on a global level. Underlying these discussions is a concern about whether the students in the state are economically competitive in a manner comparable with other states and other nations.

This competitive concern also permeates the discussions at the national level. Although policy makers at this level are concerned about differences between states, as well as types of school systems, the underlying worry is about how much American students are learning in comparison to their peers in other countries, particularly those which are our trading partners.

These information needs, which may be very different at each level from the classroom to the national level, often form the basis for assessment design. In the top-down models, assessments which meet the needs of policy makers at the state or national levels are developed and implemented. Since we presume that the data may be useful for other purposes, we try to convince others such as building principals, teachers, and parents that the information will meet their needs as well.

An emerging alternative to this is to build the assessment system which teachers, parents and students need and presume that the users at the district, state and national levels can have their questions answered just by aggregating the types of assessments used at the classroom levels.

Assessment Formats

During the past several years, as the nation has become increasingly concerned about the skills that the nation's students possess or should acquire, an essential element of the discussion of the standards needed by students and schools is the formats for assessment used in this country. Unlike other countries, the United States relies almost exclusively on the multiple-choice test as the means of assessing student performance. Critics indicate that such assessments tend to stress basic knowledge types of skills and encourage teachers to stress memorization of content, rather than the use of content to solve meaningful problems.

A variety of content-area groups are currently re-examining what they view as important and how schools should be teaching these outcomes. A common

element across these subject-matter groups is the de-emphasis of content knowledge and an emerging emphasis on application and use of the content, even to the point of conscious decisions not to teach certain content. It is this growing shift in emphasis in student outcomes which leads some at the national, state and local levels to emphasize new means of assessing student performance. Emerging techniques, such as the use of portfolios, projects, exhibitions, demonstrations, individual performance assessments, group performance assessments, hands-on assessments, and so forth are increasingly viewed as the more appropriate, more-directly-tied-to-instruction techniques for assessing students. It is ironic that these techniques, plus others (such as use of anecdotal records, observation, structured interviews and others) which are widely viewed as more suited for classroom-level assessment, are emerging as potential strategies to be used in large-scale assessment programs.

Yet, in recent years, questions have been raised about the feasibility of using such innovative assessment strategies on a widescale basis. Issues of assessment time, generalizability, quality and breadth of resultant information, and costs have emerged as major impediments to the adoption of performance assessment in many large-scale assessment programs. Policy makers and others view these instructional-related assessment strategies as the ones to use for assessment programs tied to instructional improvement.

Each of these types of assessment may be more useful or may be more feasible (due to cost, time, and the availability of the information) at different levels of the educational system. Figure 2 illustrates this point.

Figure 2
Effective Use of Assessment Strategies at Different Education Levels

Education Level	Types of Assessment Exercises					
	Sel'ed Resp	ShortAnswer	ExtendResp	Perform. Events	Perform. Tasks	Portfolio
National	XXX	XX	X			
State	XXX	XX	XX	X		X
District	XXX	XXX	XX	XX	X	XX
School	XX	XXX	XXX	XX	XX	XXX
Classroom	XX	XXX	XXX	XXX	XXX	XXX

Note: Each X indicates the levels at which each type of assessment might be used most efficiently at each level. Multiple X's indicate the strength of the utility/practicality of the assessment method at each level.

More information about assessment formats is provided in Attachment B.

Summary

As the policy makers, assessment specialists and curriculum specialists look to

design the rational and comprehensive assessment system needed by this country, important elements of this discussion include what skills are most critical, how these will be learned and what assessments will most inform and lead students to this learning? What are the assessment designs most appropriate for the states and for local districts and should these designs focus on assessment procedures traditionally used in large-scale programs, assessment strategies which have been used predominantly at the classroom and school level, or both?

Chapter Three

An Example of a Coordinated Assessment System

Each of the forty-five states that have some form of large-scale assessment program have a different configuration of grades and subject areas assessed, use different forms of assessment, and in some cases, use multiple forms of assessment. Therefore, the manner in which each state could develop a coordinated assessment system would undoubtedly look different. To illustrate what such a system would look like, the following is presented as an example of how such a system could be developed; it is not intended to serve as a model coordinated assessment system. Note: the terms used below are defined in the next chapters.

A. The state, working through state coordination teams, develops a set of content standards in several content areas: mathematics, reading/writing, science, social studies (with particular emphasis on history, geography, and civics), health education, physical education, the arts, and world languages. The coordination teams help to assure local district input into the development of the state content standards. Most school districts adopt the state standards as their own.

B. In each area, the state coordination team develops an assessment blueprint that describes the manner in which the content standards are to be assessed, both at the state, local district, and classroom levels. The assessment blueprint is also developed with local school districts so that the types of assessments that are created will be effective at both the local and state levels, and will provide a coordinated picture of student achievement.

C. The state selects the areas of mathematics, reading/writing, and science for statewide assessment. The assessments are to be administered in grades 4, 8, and 12. The purpose of the assessments is primarily to hold schools accountable for student performance. The results are reported to parents, teachers, schools, and school districts shortly following the assessments, which are conducted in the fall.

D. For each area in which the state has created content standards, the state coordination team also creates performance standards. These performance standards are created so that the assessments at the local and state levels can be used to judge the performance of students and schools.

E. For each area in which the state has created content standards, the state coordination team also develops a professional development program (the training strategies and the materials) needed to assure that all local educators are able to address the content standards to help assure that students achieve a high levels.

F. The state creates the assessments that will be used for the grade 4, 8, and 12 assessment program to be used statewide. The state coordination team oversees the work of the Department and the assessment contractor to assure that the state assessments match the content standards and fulfill the state purposes of the overall assessment system.

G. The state creates assessments (portfolio assessments, performance events, performance tasks, plus more conventional selected-response and open-ended assessments) for use at "off-grades," throughout the school year. The purpose of these assessments is to provide information to teachers to improve the learning of individual students, as well as group information to improve the instructional program at the school and classroom levels.

Although these assessments are to be used only by local school districts, the state takes the responsibility to see that the assessments are created, validated and distributed across the state. As part of this process, the state administers the assessments to a sample of students statewide at each grade level K-12, develops scoring rubrics and training materials for each open-ended or performance measure, and prepares the materials for distribution to school districts.

H. Once the assessments are created, they are tried out in a representative set of classrooms around the state. The results of the assessment are used in several ways: to refine the assessments themselves, to refine the assessment administration directions, to revise and expand the scoring rubrics using actual samples of student work, and the tentative performance standards are revised.

I. The state provides on-going information and professional development opportunities to all local school districts. The assessment information collected by classroom teachers is summarized at the building level. In addition, the district and school summaries from the state assessment information are added to the classroom-level information to provide a more complete picture of student achievement.

Chapter Four

Developing the Coordinated Assessment System

How Should the Assessment System be Developed? Top-Down or Bottom-Up? The issue of how the assessment system should be developed, top-down from the state to local school district, or bottom-up from the school or district level to the state, should be determined. Information gathered at the classroom level can be aggregated upwards, but is it the most cost-effective instrumentation to answer the questions posed by state-level policy makers concerned about the performance of school systems within the state? The portfolio of work that would be much more effective for this teacher may not be cost effective in answering the state-level questions. Is it credible? Assessment that answers the questions of policy makers at the state level probably won't serve to effectively inform classroom teachers about the learning of individual students. Given the different information needs of persons at different levels of the educational system, plus the variety of reasons why assessment takes place and the various tools available for assessment, is it any wonder that one instrument imposed at all levels will not meet all the needs and purposes for assessment?

What is needed, then, is a top-down and bottom-up approach to assessment development, with parties from each level involved in the assessment design as mentioned earlier. Since within the ideal assessment system, some data will be collected at the national and state levels, some information at the state and school district levels, some at the school district and building levels, some at the school and classroom levels, some at the classroom and student levels and some only at the student level, it is important that representatives of the different levels in the assessment system be involved in the assessment system design process.

The ideal coordinated assessment system is one, then, that has identified the various audiences for assessment information, the types of information that each audience needs (keeping in mind that each audience may have the need for more than one type of information), the type of assessment instruments that best meets the assessment needs, the impact of the use of the instrument on the educational system, and the levels at which to use the assessment instruments.

The discussion of a coordinated, comprehensive assessment system, as described in this paper, has been somewhat theoretical up to this point. In this section, the process of developing the standards and assessments in a coordinated manner is described. The implication that the reader might draw from this discussion is that unless this process is followed precisely, coordination of state and local assessments is impossible. This is not the intent of this paper. Nor is it the intent of this paper to imply that current efforts to reform teaching or assessment are not adequate.

Clearly, states and local districts already have testing or assessment programs in place, and many have created standards or frameworks to guide instruction or assessment. There are instances where local districts have built standards to amplify and extend state standards and have constructed assessments to complement those used at the state level. In this fashion, useful coordination has been achieved, although it did not occur simultaneously. Other states have

delegated the development of standards or assessments to local districts, and districts are well under way in creating the rigorous standards and assessments needed to improve education. States and local districts that have carried out these activities should receive recognition for their efforts.

The ideas presented in this paper are meant to suggest ways in which states and local school districts could *better* coordinate their standards and assessment efforts. It does not imply that all assessments developed without coordinated effort are inappropriate nor that such assessments need to be scrapped. By presenting the information in a more theoretical point of view, it is hoped that as such assessment systems are revised in the future, some of the ideas presented here can be implemented.

Process of Coordination

To develop a coordinated, comprehensive assessment system, there are several steps that must be accomplished. The process by which these are accomplished is important. For the system to be coordinated across levels of the educational system, the individuals responsible for education at the different levels must be adequately represented in the decision-making process to be described below. This means that policymakers, curriculum and assessment experts, and practitioners from each level will need to be actively involved in each step of the development process.

The first step in the process is to determine at which levels coordination is desired. Is this to be state and local district coordination, or are school buildings or national efforts to be coordinated as well? Will the state take the lead in developing standards and assessments with local districts expected to expand upon these for local use, or are local districts expected to take the lead and the state serve to coordinate the standards and assessments developed within individual school districts? If the state takes the lead, school districts will also have to attend to inclusive procedures for the development of standards and assessments at the district and school levels.

The second step is to establish teams that will encourage the types of buy-in that this process is set up to encourage. While states typically include local educators on teams that develop standards or assessments, the inclusion of individual teachers or curriculum specialists in these efforts does not encourage the school districts or schools they are employed by to value the state frameworks or assessments they help to create. Instead, a process of inclusion of teams from districts and schools (if school-level coordination is desired) is needed. However, there are usually so many school districts within a state that any team with even one person per school district would be overwhelmingly large in size. How could such a team be set and function effectively? While this process describes what states should do, local school districts will also need to attend to the manner in which the district organizes to include school-level representation on district teams.

One way to assure representation, yet keep the working groups' sizes manageable, is to use representative committees, and to encourage iterative reviews carried out by the representatives at the state level and the individuals they represent at the district level. At the state level, a state coordination team is needed, with representatives drawn either directly from individual school

districts (perhaps more appropriate for the largest school districts in the state), or selected to represent groups of school districts (selected either by size of school district or region of the state).

At the district level, a district coordination team should be established. For smaller districts, such coordination teams could be carried out across multiple school districts. For example, it is not necessary to have every suburban district represented on a state committee to know what such school districts think an assessment framework should contain. A few representatives to the state committee, selected by the suburban districts, could represent the suburban districts, particularly if given the opportunity and encouragement to share drafts with all of the districts that they represent. If school-level coordination is desired, then the district-level representatives could be encouraged to share the drafts with school-level representatives. In either case, feedback from the districts or schools could be brought back to the state coordination team from the local level.

The model described above will work whether it is expected for the state to take the lead and local districts are to coordinate their efforts with the state or local districts take the lead (working independently of one another) and the states serves to coordinate these independent developmental projects. The difference will be in the flow of information (from the state to local districts, with feedback returned to the state, or from local school districts to the state, with feedback flowing back to the local districts). When attempting to retrofit locally developed systems with one another and with the state, the flow of information may be in both directions. The suggestions that follow, while describing an ideal system, can also serve as a model for these other two models for development of assessment systems.

Once documents are completed in final draft form, they could be transmitted to all school districts (which, in turn, could be encouraged to copy them and send them to all school buildings) for a formal review by district-level (and school-level) teams. The result will be comments from these teams that can be used in reshaping the documents being examined, be they assessment frameworks, assessment blueprints, draft assessments, and so forth. In the case of secure materials, such reviews could be carried through regional meetings run by the state for the local districts in each region.

If this process is carried out in a thoughtful manner, and the resulting ideas are used to re-draft the materials, then it should be possible for coordination to begin to occur, since educators at each level will have the opportunity to determine similarity of assessment purposes, design, and instrumentation, and to provide feedback to one another to reduce redundancy and increase the collection of complementary information.

Developmental Steps

If the mechanism for coordination can be implemented, how will the coordinated assessment system be developed? The process may not be too dissimilar from that to be used in the development of a statewide assessment system. For example, first the content standards may be developed. Second, the assessment blueprint could be written, including assessment purpose(s), types of assessments to be used, and the development, review, tryout and revision

process to be implemented. Finally, the performance standards might be written to include how individual student and system (at the school, district, and state levels) performance will be determined. In other models, the performance standards might be developed along with the content standards, with the assessment following both of these (and the performance standards "verified" once the assessment is piloted). In any case, each of these activities needs to be carried out in a coordinated manner, regardless of which is developed first, second, or last.

A. State and Local Content Standards Coordination

The development of a coordinated, comprehensive assessment system, combining assessments at the state and local levels, or the retrofitting of existing assessments and standards at the state and local levels, begins with the development of a coordinated set of content standards that is suitable for use at both the state and local levels (or searching for the commonalties among existing sets of standards). At the state level, content standards serve the purposes of setting an overall direction for education within the state, as well as the framework for holding schools accountable for student performance. At the local level, content standards serve as the framework for curriculum and instruction that should be implemented at the classroom level. Can the two purposes be met with one set of standards?

One set of standards could work for both the state and local needs. Even in states where local districts are expected to develop standards that meet or exceed those of the state, it may be useful and necessary for the state to create a set of standards to be used for district-to-district coordination. Alternatively, one set of local districts standards might serve as the "standard" set of content standards around which to seek coordination, or the state and local districts could develop a consensus framework drawn from common standards at both levels. In any case, it is possible to develop a set of standards with different levels of specificity. Typically, at the state level, a set of content standards contains relatively few standards that are written at a more general level. At the district level, each of these general standards may be defined with several sub-skills or benchmark which is useful in curriculum planning. At the classroom level, each may be further defined with additional sub-skills for use in instruction. Hence, it is feasible to develop an assessment program drawing on the overall standards document, while creating curriculum and instructional frameworks, based on the same overall set of standards, but which is directly based on the lower levels of sub-skills. In addition, the bases for assessment at the different levels could be the different levels of definition of skills contained within the same set of content standards.

How could such a set of standards be developed? The process of development would need to include local school districts in the fashion described earlier that leads to "buy in" to the resulting standards (regardless of which level takes the lead). As the development process unfolds, the state-local committees can serve as a two-way communication process, providing input from local districts into the format and content of the standards, as well as serving as an on-going reaction panel for work under development. Once the final draft of the standards has been completed at the state level, two types of reviews could be invited. First, the draft content standards could be shared with professional organizations for <u>organizational</u> input to them. Second, the draft standards

could be shared with all or a representative sample of school districts within the state. This could serve to receive district input about the standards (and, secondarily, build even stronger commitment of the districts to the standards).

An additional way to strengthen district commitment to the overall content standards is for the state to provide incentive monies for local school districts to construct instructional programs based on the content standards. By funding a group of local school districts, carefully selected to represent the state geographically and by size and type, the state can further encourage local school districts to use the content standards as the basis for instruction within the school district. These efforts can be further expanded by using the same model for the development of instructional materials, professional development for educators, and public engagement materials, just to name a few. The goal is to use the coordinated set of content standards to develop a set of resources for local school districts, and to provide incentives for local districts to use these.

Where states have already created their content standards at the state level, local districts can take the standards and amplify and extend these standards for use at the local level. While states may develop standards that are few in number, local districts may need to expand upon these to provide a curriculum framework for guiding instruction (and assessment) at every grade level. Districts can do this work either alone or working with other districts.

B. Development of the Assessment Blueprint

The typical assessment blueprint serves several purposes. The first of these is to serve as a living, working document that describes how the assessment program was developed, what types of measures will be used, how these assessments are to be administered and scored, and how the results will be reported. The reporting should include what performance standards for student or system performance are to be created, and eventually, the definitions of each level of performance. To aid the reader of the document, the blueprint will eventually contain sample exercises (of each type to be used in the program), extensive exercise specifications and criteria, and sample report formats. At the outset of the development process, the document may contain only outlines of these sections, to be filled in as the assessment development process unfolds.

In cases where the state and local school districts seek to coordinate existing standards and assessments, it still may be useful to construct an assessment blueprint with the elements discussed below, since this will serve to help the participants in the process look for commonalties in assessment purpose, format, reporting, and so on. This can also serve as the basis for future development work that increase the likelihood of coordination from the start.

The first decision that the state coordination team needs to make is to determine the purposes for assessment. Actually, at some levels (such as the state level), the decision may have already been made in the legislation that enabled the program. In this case, it may be more important for the state coordination team.

1. Purposes for Assessment

It is important to consider what are the purposes of student assessment at the

student, school, district, state, and national levels. Attachment A provides a list of common purposes for assessment. Too often, the purposes of the assessment system are not well thought out, particularly in advance of the development of the instruments that will be used in the assessment system. More often than not, the program is implemented for one purpose, using instruments that are well-suited for that purpose, and then additional purposes or uses may be added to the program, where the assessment formats and purpose may fit less well.

For example, while an assessment program may be implemented to hold schools accountable, the individual student results it provides may be thought by some to provide the information that teachers need to improve instruction. However, the pressures of accountability may distort the instructional value of the individual student results. A program efficient for one purpose may not be effective for another. The question of conflicts among these purposes is rarely thought through. Therefore, it is important that policymakers who are creating the assessment system think through and prioritize the assessment purposes, so that the instruments can be matched to these purposes.

2. Formats for Student Assessment

Next, the types of assessments must be selected. As described in Attachment B, there are a wide variety of assessments to select from. Until recently, most assessment programs used primarily multiple-choice tests, with essays used occasionally where needed, such as in writing examinations. Although programs such as the National Assessment of Educational Progress pioneered the use of a broader array of assessment formats, much of that innovation was dropped for budget reasons twenty years ago and, hence, does not serve to model other possible designs.

Educators will also need to choose whether to use criterion-referenced or norm-referenced measures. Criterion-referenced assessments describe whether or not students have certain knowledge and skills, while norm-references assessments express student performance in comparison to a norm group. Examples of each type of assessment can be obtained commercially, or they may be developed within each state. It is important not to confuse the manner in which results are reported (relative to norms or standards) with the types of assessment exercises such assessments typically contain.

Because the nature of the skills that educators consider to be important is changing, the manner in which the skills are assessed may need to change as well. Alternatives are increasingly being used in large-scale assessment programs. Such alternatives as the use of essays or extended response formats in assessing entire areas such as mathematics or reading, the use of interview questions, hands-on performance exercises, group performance exercises which are scored both for end result and group process are some of the "on-demand" types of exercises now finding their way into large-scale assessment programs. Some programs are now using a few of these types of exercises, while several state programs are now based solely on these alternative programs.

In addition to these on-demand types of exercises, other assessment formats which break out of the time constraints and allow students to develop a response over time are currently being used in a few large-scale programs. Such assessment formats include the use of portfolios, demonstrations, exhibitions,

and projects. As the large-scale assessment program is being designed, both the purpose(s) for assessment, as discussed in the previous section, and the types of assessment formats to be used, must be considered. The selection of the measures to be used should be based primarily on which strategies are needed to effectively, accurately, and efficiently assess the content standards. By dividing up the assessment load among levels, it is possible to use assessments that are more costly and time-consuming where they will do the most instructional/ learning good – at the classroom level, while using measures at the district and state levels that are supportive and complementary.

C. Developing Performance Standards

Performance standards determine how well students or schools need to perform in order to reach different pre-defined levels of performance.

Once the content standards to be used across levels are developed (or the standards developed at different levels are conceptually mapped together), and the assessment design has been determined, the next step may be to determine the anticipated desired levels of student and school performance. The performance standard may also be developed in concert with the content standards, or in other systems, the performance standards may not be developed until student information has already been collected. One advantage of setting performance standards in advance of creating the assessments, at least in draft form, is that the performance standards can help determine at what levels the assessments should be developed, as well as the content of specific assessment exercises. Once the assessments have been created and validated, these verbal definitions can be refined using the assessment information collected during tryouts.

The first step is to review the range of student performance(s) the content standards and/or assessments may elicit. Then, decide how many levels of student performance are desired or needed. The basic decision, of course, is "pass-fail." However, such gross levels of description are generally too broad for most intended uses of the information. Therefore, generally there is a more advanced level of passing and another more basic level of not passing. For IASA Title I purposes, where initially many students are expected to do poorly on the assessments, it may be desirable to have three or more levels of not passing. This will allow smaller increments of improved student performance to be noted in future reports of assessment results.

The next step typically is to further define what students should know and be able to do at each performance level. These descriptions should be as specific and as behavioral as possible, so that persons unfamiliar with the content standards or the assessment to be used can understand them without reference to either what it means to be "proficient," "advanced," or other labels. By being stated in specific terms, this means that the developers of the assessment will have a good idea of the types of assessment exercises that must be created. Typically, these descriptions are used to guide the development of the assessment blueprint, and more importantly, are used to assure that there are sufficient numbers of exercises at or about each performance level in order to classify students accurately. If student work is available, through informal or formal field tests, the student work can be used both in the development of the performance standards, or the "validation" of the levels and standards once the assessments

are formally field-tested on students.

Next, decisions need to be made about what levels and procedures are to be followed in categorizing school-level performance. How will group performance in any year be described? How will improvement over time be determined? How much change and of what kinds is needed to describe a school as "improving?" These are issues that should be determined at this point.

Additionally, will the determination of performance be restricted to the assessment created for each level of the educational system, or will the assessment data collected at different levels with multiple assessments be used to make the overall determination of level of performance? With a coordinated assessment designed to meet the assessment/evaluation needs of IASA Title I, it may be possible to establish cross-level performance indicators that draw from information collected at the state and local levels. For example, judgements of the level of student performance could be based on the performance of students on the state assessment program (comprised of selected-response and open-ended exercises), plus the performance of students on portfolios collected and rated by classroom teachers. This would require, of course, substantial coordination not only of the content standards used at different levels, but also the assessment strategies used at each level to assure that the different measures used really measure the same sorts of standards. This is not an activity that has been carried out in the past, in large part due to the lack of coordinated assessment systems spanning state and local levels; however, this is an area where fruitful developmental research is needed.

The next step is to develop the assessments to be used following the assessment blueprint and based on the content standards. During the assessment development process, data from students will be collected during the tryout phase. This data, or the data from the initial live use of the assessment instruments, should be used to establish the final performance standards for both students and schools. There are several established procedures for setting student performance standards, such as the modified Angoff method or the contrasting groups method, or other newly-emerging strategies for setting cut scores. One or more of these strategies can be used to set the final levels of desired performance, as well as to revise the descriptions of these levels to match the actual performance standards.

Finally, the anticipated school performance standards will need to be reviewed for realism once either tryout or live data is available. Do students perform as was anticipated, or were the *a priori* judgements too high or too low? Policymakers, parents, business leaders, and others can help educators judge the realism of the standards.

Chapter Five

Issues in Developing Coordinated Assessment Systems

There are a number of issues inherent in developing coordinated assessment systems. Some of these are ones inherent in any attempt to develop an assessment system, while others are important as coordination is being sought. Regardless of where a state and its local districts are in developing a coordinated assessment system, these issues will be important to consider.

Alignment of Standards As the development process unfolds, the state coordination committee will need to wrestle with the issue of what constitutes "alignment of content standards." The question of whether a particular local district's content standards are aligned with the state's content standards, as well as whether the content standards developed in one school district will match those developed in another district, will arise. Both are important issues to consider. One strategy is to use the state's standards for examining the match to local districts' standards, encouraging each district to maintain the standards unique to the district.

Even if the standards developed at the district level or state level are developed at the same time and even if there is overlap in the committees that develop the standards, there may be different interpretations of similar standards, and there may well be different levels of specificity of the standards. One way to respond to both questions is to conceptually map the different sets of standards together.

How does conceptual mapping occur? One strategy is to have the developers of the different sets of standards examine first their structure and their content at comparable levels of specificity.

- Does one framework contain more specific examples of the standards contained in the other set?
- Have different districts developed different sets of more specific examples of the state standards?
- Are the sets of more specific examples of state standards related to one another?
- Are there important differences that should maintained if one were developing an overall set of standards containing all of the important standards, or are apparent differences simply different ways of describing the same thing?

By carefully considering questions such as these, the group that is examining the different sets of standards can compile one set of standards that represents a complete set of the different standards drawn from state and local districts that contains all of the important standards valued at the different levels of the system. By doing this exercise, the group carrying out this conceptual mapping can determine the extent of similarity or differences among the different sets of content standards. This will serve to provide an in-depth answer to the question of whether different sets of standards are aligned.

Alignment of Assessments In a comprehensive assessment system, some attention will need to be paid to whether the different assessment formats to be used will actually measure the same standards. For example, if the state is assessing state standards using a mixed assessment model comprised of selected-response and open-ended exercises, and the district is conducting assessment using open-ended exercises and performance event-type assessments, and classrooms teachers are assessing student performance using a combination of selected-response, open-ended, and portfolio assessments, do the different assessments used at the various levels measure the same standards? Assuming that the redundancy in assessment types was planned, will the assessment results from various levels relate to one another?

Each assessment will need to be compared to the content standards that it is measuring to determine whether the assessments do match the content standards. There are several key questions that assessment system designers need answer:

- Does each assessment measure an important aspect of one or more content standard?

  This question poses the issue of whether the assessment covers an important element of the content standards. Quality assessments measure the content standards. Is there one or more standards for each assessment exercise?

- Are the content standards well measured by the assessment (that is, does the assessment assess the breadth and depth of the content standards)?

  This question is the reverse of the first one. The alignment of assessment to standards needs to be examined from the perspective of the standards as well. How well are the content standards assessed by the assessment(s) used?

- Does the format(s) used in the assessment convey the sense of the importance of the balance between content and skills?

  The nature of the assessments used conveys a sense of the meaning of the standards. For example, if the standards indicate that students are to design innovative approaches to studying a scientific concept, yet students are simply asked to select answers to multiple-choice questions regarding science content, the format of the assessment may not be well aligned with the standards being assessed.

- Will the assessment reports convey the sense of what is important in the content standards by providing the emphasis on content and skills contained in the content standards?

  Regardless of the types of assessment exercises being used, how will the reports of results match the types of standards assessed? Will student performance on the most important standards be reported? Will it carry greater weight than student performance on other standards?

Can One Assessment Meet Multiple Purposes? The issue of whether it is <u>really</u> necessary to have a different test for every assessment purpose is one that always

arises, since each assessment costs time and money. The evidence suggests that the answer is that multiple assessments are needed, since trying to fulfill too many purposes may lead to fulfilling none of them well.

Do Some Assessments Work Better for Some Assessment Purposes? Some assessment purposes are best achieved by particular types of assessment. The match of assessment type to purpose is important.

Do Local Schools Really Desire Coordination? Too often, local school districts have either waited for the state to develop its system, or have tried to minimize its impact on instruction in the classroom. By working together, it is hoped that states and local districts will establish a pattern of trust needed for coordination of standards and assessments to occur. In the short-run and in some cases, it may be difficult to establish this trust and working relationship. The state can still offer to work together and provide resources and incentives for local schools to use, hoping that the materials and resources will be so compelling that local districts will choose to work with the state.

Does the State Desire Coordination? The answer to this question is not as simple as it may seem. While some at the state level may strongly desire to provide technical assistance to local districts and to work on assessments for local districts that coordinate with state assessments, state-level personnel may not be given the resources or the opportunity/mandate to do so. In some cases, working with local schools might be viewed suspiciously. It is important to clarify for persons who do not understand the reasons for such work why it is important to work together. Such coordinated work should not be viewed as diminishing the value of accountability assessments.

Are There Professional Development Implications of Coordinated Assessment Systems? For coordination in the development and use of assessment systems to truly occur, all participants in the system (policymakers, administrators, teachers, parents, and students) will need information about assessment suitable to their background and interests. This implies a substantial training on assessment. Given current relatively low levels of assessment literacy, this will mean a concerted effort to improve assessment literacy among educators and non-educators, and at all levels of the educational system.

Summary

These are just some of the important questions that the assessment system designers will want to wrestle as they develop and implement an assessment system on a coordinated basis to assure that information collected at different levels is useful in creating a mosaic of information across levels.

Chapter Six

Summary

A variety of reasons and purposes for assessment exist, and there are a variety of audiences for assessment information at the student, school, district, state, and national levels. There are also many assessment strategies that have been used, each with advantages and disadvantages. Designing the ideal assessment program, the one that meets the needs of persons at these various levels, consists of matching the purposes with appropriate types of assessment instrumentation. The full variety of measures should be considered when designing the system because some types are better suited for some purposes than others. An illustration of the school improvement team was used to convey that there is a strong advantage in using multiple measures at the level of the school building to collect information on the success of the school in meeting the objectives that the staff has developed.

Generalizing from this, the ideal assessment system would attempt as much as possible to capture information from the bottom up while adding measures from the top down, so that users at all levels may have more than one approach to assessment to use when reporting the status of the system at that level. This approach has the advantage of demonstrating that different approaches to assessment may yield different results and that these differences help to inform the debate on what skills students should be able to perform and whether or not students are currently able to demonstrate performance on these.

If such a coordinated assessment system can be created, the advantage for IASA Title I is that more information about the performance of students will be more readily available, and this information will serve to assure that Title I students are being held to the same standards as all other students. It can also help assure that the assessments used for program evaluation purposes do not serve to distort the educational program of students by emphasizing only lesser important aspects of the standards being assessed. This alone will help to correct a major problem with past evaluation designs for compensatory education. A coordinated assessment design will help to assure that assessments used at different levels are well-suited to accomplish the purposes for which they are being used, without being used in ways for which they are not designed nor well-suited.

When the steps outlined in this paper are followed, the end product will be a cross-level coordinated assessment system, using different modes of assessment to assess a common or conceptually-linked set of content standards, and reported via performance standards established for the assessments at each level or across levels. By using multiple measures to accomplish different assessment purposes, and reducing the amount of redundant information as well as the mixed messages sent by uncoordinated assessments based on different standards, result will be a more coherent picture of student performance at the individual and group levels.

Attachment A

Purposes for Assessment

The major purposes for assessment can be summarized as follows:

Monitoring
- Student Level: Provide periodic measurements of student progress in order to determine the educational "growth" of a student from year to year.
- System Performance: Provide a periodic measure of the performance of groups of students to track performance over time.

Information/Accountability
- Parents and Students: Inform parents/student about student performance so as to encourage student or teachers to improve performance.
- Public: Provide the public with information about the performance of groups of students so as to encourage schools to improve the system.

Improving Student Performance
- Student Level: Provide data to teachers and students that encourages instruction geared to the needs of individual students to help them achieve at high levels.
- System Level: Provide information to educators on groups of students, such as the school level, which can be used to review current instructional strategies and materials at one or more grade levels and be used to make improvements where needed.

Allocation of Resources
- Human: Use information to determine where additional staff are needed.
- Financial: Determine where financial resources are most needed or should be used.

Selection/Placement of Students
- Selection: Help determine the eligibility of students for various educational programs or services
- Placement: Determine the program or service most appropriate for the instructional level of the student.

Certification
- Student Level: Provide a means of determining the competence level of individual students.
- Program Level : Provide data to certify the adequacy of an educational program, such as advanced placement courses.
- System: Provide data to certify the acceptability of an educational system, such as in accreditation programs.

Program Evaluation
- Provide the information needed to determine the effectiveness of an educational program or intervention.

Attachment B

Formats for Assessment

Some formats for assessment are listed and defined below:

On-Demand Assessments

- Selected-response (multiple-choice) exercises: In these types of exercises, students select one or more answers from a list of suggested responses. These exercises have the advantage of not taking much time to complete, which may make them well-suited for assessments of a broad set of content. However, because students select a response, it is more difficult to assess student thinking with such items.

- Short-answer, open-ended: In these types of exercises, students write in an answer to a question. The response is typically a phrase, a sentence, or a quick drawing or sketch. Response time is generally limited to ten minutes or less. Less guessing is involved than in selected-response exercises, but these exercises do not tap much student thinking, either.

- Extended-response, open-ended: These exercises require students to compose a response that may be several pages in length, and require fifteen minutes or more for a complete response. They require much thought on the part of the student. Still, this is only first-draft student work, so it may not represent what students could do given additional time and encouragement in which to compose a final draft.

- Individually-administered interview: In these exercises, an assessment administrator administers the exercise individually to students. This format permits the interviewer to ask each student questions about the topic being assessed, which may be needed in order to judge the performance of the student on other parts of the exercise. Has the advantage of tapping important skills often desired of students, although this is quite expensive both to administer and to score.

- Individually-administered performance events: These are exercises that are completed by individual students within a class period, and involve some type of performance on the part of the student. This may be because the exercise requires special equipment (as in a science experiment), requires an individual student to perform (such as playing a musical instrument), or because we may wish to observe the process that a student used to respond to the question (such as in a mathematics problem-solving activity).

- Group-administered performance events: These are exercises that groups of students respond to. These may be existing groups (e.g., a band or orchestra) or groups made up just for assessment purposes (a group of six students assessed for teamwork skills). The students typically perform in some fashion and the group interaction and/or performance is scored as a whole (and, perhaps, individually).

Extended-Time Assessments

- Individual performance tasks: These exercises are ones on which students

may work for several days, weeks, or months to produce an individual student response. This could be, for example, a science experiment (e.g., design and plan a garden and make observations about it over time) or other performance that students will need time to complete.

- Group performance tasks: These exercises are ones on which groups of students may work for several days, weeks, or months to produce either a group and/or individual student responses. This could be, for example, a health education task (e.g., design a school lunch menu for a month that is nutritious, affordable and that would appeal to students) or other performance that students will need time to complete.

- Portfolio assessment: There are several reasons why portfolios of student work may be kept. First, it can serve to document the changes that students are making in a work (e.g., different iterations of an essay). Second, they can use the portfolio to assemble a collection of their best or most polished pieces (e.g., a collection of musical performances). Third, and perhaps most importantly, students can use their portfolio to document their ability to achieve important outcomes, such as those contained in state or national content standards. In this case, the portfolio provides the evidence that the student needs to demonstrate his or her competence on the standards, with the demonstration being provided by a persuasive piece that the student provides to the scorer.

- Observations: Other important types of information that teachers can collect over time come from structured and unstructured observations of students. Structured observations, for example, might be made in a pre-arranged classroom set-up in which students are given several choices of free-time activities and observed as to which ones they engage in and for how long. Unstructured observations are the events that occur within the day-to-day classroom that teachers may wish to record for the future. For example, if a student is having difficulty in mathematics, the teacher may observe that the student is not listening during instruction and therefore not picking up the knowledge needed to take tests.

- Anecdotal records: There are other sources of information about students, such as notes from other educational professionals, parents, and others. These records may also provide useful information about individual students.

Figure 1
Assessment Purposes at Different Levels of Education

Education Level	Purposes for Assessment: Monitor	Account.	Improve Student Perf.	Allocate Resources	Selection/ Placement	Certification	Program Evaluatio
National	XXX	X		X		X	
State	XXX	XXX	X	XX		XXX	X
District	XX	XXX	XX	XX	XX	XX	XX
School	X	XX	XXX	X	XXX	X	XX
Classroom		X	XXX		XX	X	X
Student		X	XXX		X	X	X

Note: Each X indicates the levels at which each type of assessment might be used most efficiently at each level.

Figure 2
Effective Use of Assessment Strategies at Different Education Levels

Education Level	Types of Assessment Exercises: Sel'ed Resp	ShortAnswer	ExtendResp	Perform. Events	Perform. Tasks	Portfolio
National	XXX	XX	X			
State	XXX	XX	XX	X		X
District	XXX	XXX	XX	XX	X	XX
School	XX	XXX	XXX	XX	XX	XXX
Classroom	XX	XXX	XXX	XXX	XXX	XXX

Note: Each X indicates the levels at which each type of assessment might be used most efficiently at each level.